DECISION SUPPORT IN A SERVICE FIRM

Gallo Mosè(a), Murino Teresa(b), Romano Elpidio(c), Fasano Massimo

[(a) (b) (c)] Department of Material Engineering and Materials
University of Naples "Federico II"
P.le Tecchio, 80125 Napoli

[(a)]mose.gallo@unina.it, [(b)]murino@unina.it, [(c)]elromano@unina.it, maxfasano7@hotmail.it

ABSTRACT
The goal of this study is to produce a decision support simulation model with discrete event logic applied to a health spa service. The simulation is an extremely useful tool for predicting the constant changes that take place in a highly dynamic context such as that of services. In particular, this study develops a simulation model can be used to build a structure that helps predict delay and to produce a logical and rational management of queues and therefore to reduce the inefficiencies of the service.
The model was tested in a medium-sized Italian health spa (Telese thermal baths).

Keywords: discrete event simulation, service system, resource management, decision support system.

1. INTRODUCTION
The thermal spa studied is a service company, that is, an operating system consisting of a combination of infrastructure, facilities, equipment, systems and personnel that collectively deliver a service to the customer.

Operational efficiency has a direct impact on customer satisfaction and also on the financial performance of the company.

In particular, this paper is aimed at the study of issues related to the carrying out of thermal treatments, with the aim of streamlining the health spa service system, focusing attention on the main customer of the spa.

The slow expansion of capacity (limited by physical dimensions that cannot always be further extended) has increasingly complicated the ability of the health spa to maintain satisfactory customer service.

The management has essentially two main objectives: customer satisfaction and cost effectiveness. On the one hand, it must seek to contain the queues, which increase the cost of customer inconvenience; on the other hand, it must consider the limited capacity available.
As a result the task of management, following a capacity planning analysis, is to adopt systems to optimize resource utilization and maximize customer satisfaction. Determining the optimal combination of resources, processes and technologies that will produce a satisfactory service for the customer is a complex problem.

Using analytical methods, customer waiting times in queues cannot be readily predicted with accuracy. The best way to take account of all the variables is undoubtedly a simulation that offers important benefits such as flexible modeling and immediacy of understanding through animations.

In the field, it was realized that a very detailed analysis of the existing process is required for understanding of the times and methods, highlighting in particular the organizational and procedural inefficiencies that affect the quality of the service, causing delays with potentially negative impacts on customer relations and therefore on the market performance of the company. We have therefore tried to offer an ad hoc methodology to limit these delays and the related costs.

2. SERVICE SYSTEM
Every production system uses resources to convert input factors to desired output factors. These transformation processes are carried out in different fields, and what differentiates them is the nature of the output. The service is "the intangible output of the production process"; its characteristics are not to be identified with the physicality and materiality that characterize the output of industrial production (tangible output). The service can therefore be defined as an exchange of activities between a subject who expresses needs, wants and aspirations and another subject who responds to these. It is therefore essential to valorize the service, as a tool to solve problems and respond to the specific needs of the customers.

Another aspect to be underlined is the lack of storage, a feature that makes the production process difficult to standardize and to schedule. This consequently means that production capacity must be immediately available to produce a service as required, preferably located at the point where the customer requests the service. Another feature of a system that provides services, is the significantly higher volatility of demand relative to manufacturing systems. This happens primarily due to lack of storage, which does not allow reserves to be accumulated during slack periods for use during periods of peak demand, making it difficult to maintain a steady level of output. Secondly, this level of variability is also caused by the "personalized" nature of the service – the delivery process is addressed to different customers who often have different needs.

Within the service sector, quality can only be observed from the perspective of customers. The service is generated in response to their needs, wants and aspirations. The performance is of high quality if the customer can enjoy the results. The organization must therefore plan the service not just according to its own needs (its problems, constraints, internal costs, and difficulties), but also to address those of the customer.

2.1. Constraints and objectives in service models

Supply differentiation is the fundamental tool to capture customer preferences. The operational features and the constraints on differentiation in service models are more varied and often more difficult to specify than those used in manufacturing.

One constraint consists of the time window related to the demand or supply of the service. There are capacity constraints related to the physical dimensions of the structures that deliver the service within the limits of the available space. Then there are constraints relating to the availability of physical resources (such as facilities and equipment) and human resources (staff). The basic problem is to optimize resource management so that wastage is reduced.

The primary goal of all manufacturing systems is to minimize production costs associated with a given level of quality, that is, the provision of a high quality product to the customer within the time limit. In service models too, one of the objectives is the minimization of the costs of service delivery. Other main objectives may be different, such as minimization of the average time of service delivery, minimization of the average customer wait time, minimization of the average distance covered by the customer in the service system, or minimization of the average number of staff required.

One of the most representative parameters of service quality is constituted by time. Just as in manufacturing companies, time is an important parameter because it is necessary that the product ordered by the customer is delivered within the time limit. In service companies, customer waiting times for service delivery must be below a tolerance limit.

A satisfactory service for the customer is therefore a service that minimizes its own inefficiencies, such as queues and customer waiting times for service delivery.

2.2. Queuing theory

Queuing theory aims to develop models for the study of waiting lines that may form in the presence of demand for a service. When the demand for a service and/or the capacity to provide that service are subject to uncertainty, this can result in temporary situations where the service provider is unable to meet customer demand immediately.

The arrival of customers is usually random and each customer requires a service with a delivery duration that also varies in each case. If a new customer arrives and the service is exhausted, he/she enters a waiting line until the facility becomes available. In the service sector, since queues are formed of people rather than tangible goods (as may be the case in manufacturing), the problem of congestion results in various hardships and inconveniences.

Queuing theory analyzes waiting lines by modeling the behavior and process of the arrival of a customer (service demand), and the modalities and process of delivery (service supply). As in any situation where supply exists in respect of a demand, it seeks to find a balance between conflicting needs:

- customers want to receive the service as soon as possible;
- the service manager must design the service system so as to maximize the level of customer satisfaction and at the same time to minimize the costs incurred in service delivery, primarily infrastructural costs and staff costs.

By means of a model (descriptive or simulative), queuing theory studies the queuing system and seeks the optimal level of resources at the minimum total cost (sum of the costs of waiting in queues and of resources), so as to balance the above objectives. In most cases, as already mentioned, the demand for the service and the duration of the delivery of the service are variable, so system sizing and effective coordination among its constituent parts are of key importance. Queuing theory is an excellent tool for planning and managing a service system for this purpose. The sizing of a service system, whether simple or complex, is determined on the basis of certain fundamental parameters, such as the average length of queues, average number of users in the system, and average customer waiting time.

Simple system types can be analyzed analytically, but when the system becomes complex, as when multiple queues are present in the system and the service involves different operations, then analytical study becomes very difficult and the only solution to estimate system performance is the use of simulation models.

The elements that define a complete service system, and therefore the phenomenon of waiting lines are:

- customer population;
- customers arrival process;
- number of servers;
- service process;
- service capacity;
- queue discipline.

The customer population is the set of potential customers arriving in the service system and leaving it after being served. The main characteristic of the population is its size, which represents the total number of different potential customers requiring a service. The behavior of customers is important: they may be patient and willing to wait (for a long time) or they may be impatient and leave after a while. For example, in call centers, customers will hang up if they have to wait too long before an operator is available, and they possibly try again after a while.

The arrival process describes the way in which customers request the service. It is defined in terms of interarrival times, i.e. the time interval between two successive arrivals. This process may be deterministic, but in general it is described by a random variable. Usually we assume that the interarrival times are independent and have a common distribution. In many practical situations, customers arrive according to a Poisson stream (i.e. exponential interarrival times). Customers may arrive one at a time, or in batches. An example of batch arrivals is the customs office at a border where travel documents of bus passengers must be checked.

The service process, on the other hand, describes the way in which each server delivers the service. It is defined in terms of service time, that is, the time required by a server to provide the service. As with the arrival process, the service process may be deterministic, but in most cases it is described by a random variable.

It is essential to define the number of servers in the service system at the time of the analysis. If there is more than one server, it is also essential to distinguish whether these work in series, if the service requires several operations to performed, or whether they work in parallel, if the service requires a single operation. Service capacity is the maximum number of users that can be present simultaneously in the system, including both customers in queues and those who are using the service. Customers, who arrive after that this capacity is saturated, are rejected.

Queue discipline describes the way in which customers are "selected" from the queue to be served, more specifically, the rule that a server uses to choose the next customer from the queue when the service of the current customer is complete. Commonly used queue disciplines are:

FIFO (“first-in first-out”): the customers are served in order of arrival;

LIFO (“last-in first-out”): the last customer to arrive is served first;

SIRO (“service in random order”): customers are served in random order;

Service based on classes with different priority, where customers with the highest priority are served first.

The costs are usually split between variable costs, which are a function of at least one of the quantities characterizing the dynamics of the system, and fixed costs, which are independent of the observed dynamics and which are generally a function solely of the physical structure of the system. In a queue there will always be present at least the variable costs associated with customer waiting time and fixed costs related to the number of servers available. The different actors involved in the system obviously consider these costs with different emphasis. Customers consider the reduction of waiting times in queues to be essential, while the manager of the service system is probably interested in achieving the maximum utilization of resources while seeking to satisfy customer needs.

3. A SIMULATION APPROACH TO THE ANALYSIS AND CRITICAL ASSESSMENT OF A PROCESS

The term simulation refers to a wide range of methods and applications that allow reproduction of the behavior of a real system, building a model with the purpose of reflecting the behavior of a physical system so that it can be studied from different perspectives and points of view.

Simulation is a very powerful tool that is widely used in the study of systems. It is typically applied in the following areas:

- design and analysis of a manufacturing process;
- design and analysis of a service process (call-centers, fast food, hospitals, banks, post offices, thermal spas, etc.);
- establishment of ordering and inventory strategies;
- design and operational implementation of transportation systems (highways, airports, ports);
- analysis of financial and/or economic systems; etc.
- Two different types of model exist:
- physical models, where there is a scale model of the system (a plastic model, flight simulator, etc.);
- abstract or logical models, where a number of assumptions are made, with all the consequent approximations regarding the structure and quantities, to reflect as far as possible the functioning of the real system. A logic model is usually represented through a program that permits queries to be made regarding its behavior and consequently that of the physical system.

After developing the model, we must interact with it, in order to analyze its behavior, and therefore (with some approximations) that of the physical system. If the model is simple enough, it is possible to solve it with a numerical / statistical / analytical approach: integral equations, linear programming, queuing theory, etc. Most systems that are modeled are rather complex and structured and so it is not possible to identify the mathematical laws that govern the system. In such cases, it is more effective use a simulation approach. In any case, the simulation is affected by uncontrollable input variables (random) which propagate in the model until they reach the exit.

3.1. Steps to perform a simulation

A simulation consists of ten steps. The first step is to formulate the problem and to plan the study. Generally, the project manager proposes a problem and, in a series of meetings in the presence of simulation analysts and experts, the following points are discussed:

general objectives of the study;
specific questions that the study should answer;
system configurations to be modelled;
software to be used;
schedule of activities of the study.

The second step consists in data collection, where information is gathered regarding operating procedures, the system layout, the data input system. The data should be made available in electronic format and organized in an appropriate database. The data should be traceable to a particular probability distribution, so that they can be included in the simulation model.

The third step is the definition, construction and representation of the conceptual model using graphical languages (such as flow charts, Idef0, Petri nets, etc.). After having defined and constructed the conceptual model, it is expressed in a programming language (Fortran, C, Matlab, etc..), or in simulation software (Arena, AutoMod, Extend, ProModel, Witness). The direct use of programming languages has the advantage of fairly low costs that lead to a fast and efficient simulation. Simulation software, on the other hand, is fast and easy to use, thus reducing programming time.

The fifth step is to create simulation tests in order to provide the data necessary for the next step. This is the validation of the simulation, in which performance measures of the real system are compared with those provided in output by the simulation model.

The seventh step is the design of experiments so as to construct an experimental plan that minimizes the number of runs for each simulation and maximizes information on critical aspects of the problem. This is followed by the carrying out of the simulations themselves, and the analysis of output data with comparison of alternative system configurations. The tenth and final step is the documentation, presentation and interpretation of results.

3.2. Discrete Event Simulation

A simulation involves the chronological representation of system states over the observation period (simulation run length), identifying those elements within the system with independent life (entities), having particular characteristics (attributes) and interacting according to particular laws when carrying out activities. The entities use the resources to carry out the activities, and the execution of one or more activities generally corresponds to a change of system state, and therefore, to the occurrence of an event.

In continuous state models, the dynamics of the system state depends on time and for this reason they are denoted as time-driven systems. In the case of discrete state systems, however, the state changes instantaneously at particular moments of time in which certain conditions associated with events are verified. Conceptually, an event can be considered as:

- a particular action (such as receiving a telephone message, the shutdown of an engine);
- the spontaneous occurrence of a condition (failure of an engine, voltage drop on a particular device);
- the occurrence of certain conditions (the number of packages in a storehouse reaches a predetermined quantity).

Assuming that a time signal (clock) is available to measure the time, there are two possible ways to generate events:

- in all multiple instants of the period of the clock there is an event e. If neither event takes place in a particular instant, we add the null event ε to the set of possible events ∑ that cause the transitions of the state of a system;
- the events occur in instants of time that are not known in advance and do not necessarily coincide with the multiples of the clock period.

In the first case, the state of the system is evaluated at each clock signal. The state may change depending on the generation of an event e or of the null event ε. There are, therefore, transitions of state synchronized with the clock signal, and the dynamics of the state of the system remains time-driven. In the second case, however, the state of the system changes at instants of time that are not known in advance, whenever an event is generated and, and transitions of state are driven by the succession of events. These systems are denoted as event-driven systems.

Summarizing we have identified two classes of dynamical systems:

- continuous state systems and time-driven systems (linear or nonlinear, time-varying or not, continuous time or discrete time);
- discrete state systems and event-driven systems (the systems of this class are called discrete event systems DES).

In the majority of problems for which a simulation approach is adopted it is necessary to manage the queues within the system. For these problems, a discrete event simulation is appropriate. In a discrete event simulation, which may be based on both continuous and discrete states, the state variables only change at discrete events (special instants), which are determined in turn by activities and delays, and not continuously (such as the volume of a gas or the level of a liquid). The basic objects of the system model are entities and resources. The former are elements of the system to be considered individually, information on the state of which is maintained during the simulation (such as passengers who arrive at the check-in waiting to be served, or patients who arrive at a hospital reception). Resources are those goods required by the entities in the course of activities and are usually elements of the system that need to be modeled individually, but which are usually available in limited quantities. It is obvious, therefore, that the entities needing a resource are blocked if that resource is not available.

An event is the instant of time when there is a significant change in the system, typically corresponding to the start or end of an activity. For example, in a queue, a possible event is the moment when the server stops serving one customer and becomes available to serve the next customer. The activity in this case is to serve the customer.

To make the simulation it is necessary to build the model, for which the following steps are required:

- identify the classes of entities;
- identify the activities that take place in the system;
- identify the relations between entities and activities.

4. HEALTH SPA SERVICE

The health spa service is complex and highly structured, characterized by various activities such as the bottling of water, traditional thermal treatments, and health and fitness services, including cosmetic activities. The connection between the thermal spas and wellness, in those destinations where it has been established, has attracted more young customers in their thirties and forties, compared to the over-fifties who characterize demand for traditional thermal treatments.

The strategic importance of the thermal spas and treatments as a specific component of tourism, was progressively established in France, Hungary and the Czech Republic in the first place and also in Slovenia, Bulgaria and Austria. In particular, with the liberalization of market, Eastern European countries have accelerated investment, offering a competitive supply system, in terms of quality and price, thanks to the collaboration of international hotel chains. Italy is the European country that has the largest number of thermal spas, thanks to its particular geological structure, rich in volcanic phenomena. 380 companies are active across a very diversified area. The regions with the highest number of thermal spas are Campania (114) and Veneto (110), followed by Emilia-Romagna (24), Tuscany (22), Lazio (18) and Lombardy (16).

4.1. Thermal activity at Telese thermal spa

The thermal spa of Telese, in the province of Benevento, has an ancient tradition of sulphurous waters, famous their numerous therapeutic applications. The thermal activity is mainly performed in the central departments of the thermal spa and partly in the decentralized departments of the associated hotel complex. In particular, thermal treatments are carried out in the central departments of the thermal spa from May to November, and in the decentralized department of the hotel for the rest of the year. The period of high flow and thus the most critical period is the month of September. The therapies performed include mud baths, warm baths, hydromassages, vaginal douches, inhalations, pulmonary ventilations, entotympanic and politzer insufflations, and massages. There are separate departments for each type of treatment and each type of client, whether private individuals or belonging to a group that has an agreement with the company. In the study we have focused only on the non-booking departments, in which there is highest criticality in terms of long queues and high waiting times in queues. The departments that are non-booking are the inhalation department, massage department, and warm bath and hydromassage department. The thermal treatments can be obtained either on prescription of the family doctor or privately. The main difference between the two types of access is that those accessing the thermal baths on the prescription of the doctor must pay only a proportion of fee, and they may be exempt altogether. With private access, the customer must pay the entire tariff determined by the company. Every citizen has the right to take a single course of treatment per year for a period of 12 days, from the Monday to the Saturday of the following week.

The company has agreements with municipalities, and also has a special agreement for lodging in the associated hotel with some organizations and national institutes.

The opening times of the departments are, in principle, from Monday to Saturday morning from 8.00 to 12.30 and from Monday to Friday afternoon from 15.30 to 17.15. During the months of high flow the company prefers to open the departments in the morning at 7.00 and in the afternoon at 14.30.

4.2. Text Paragraphs

The treatments take place in the departments of the health spa, located inside the thermal park. It has two separate entrances, a central one for groups that have an agreement with the company, and a side entrance for private customers. There are therefore two separate control stations at the entrance.

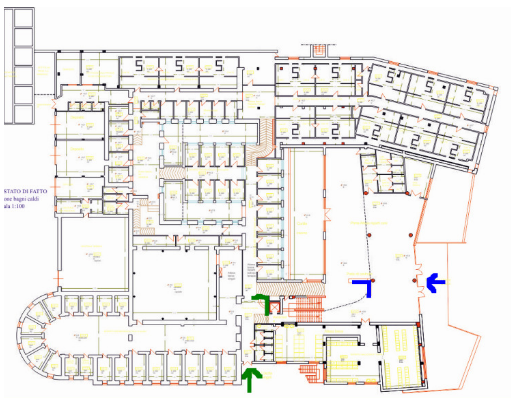

Figure 1: two separate entrances with two separate control stations for groups and private customers

We have focused only on non-reservation departments: the inhalation department, massage department, and warm bath and hydromassage department. On the lower floor is the inhalation department dedicated to the groups, on the upper floor is the inhalation department dedicated to private customers. Both have 27 aerosol appliances, 27 inhalation appliances, and 10 appliances for nasal showers and there are 3 operators in each.

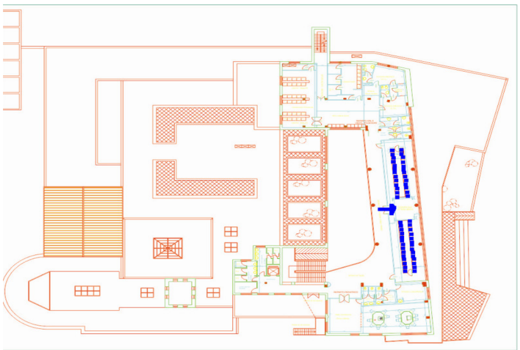

Figure 2: upstairs the inhalation department dedicated to the private customers

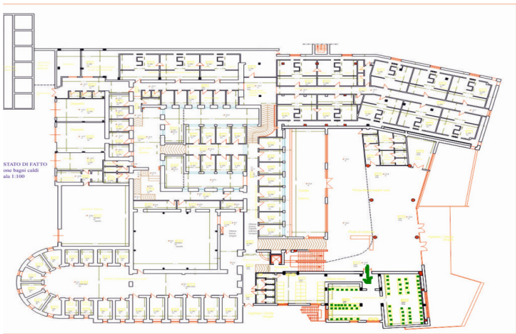

Figure 3: Lower floor inhalation department dedicated to groups

The massage department consists of 5 massage cabins used out of a potential availability of 8 cabins, with 5 operators. There is a single department for the two types of customer and it is located downstairs.

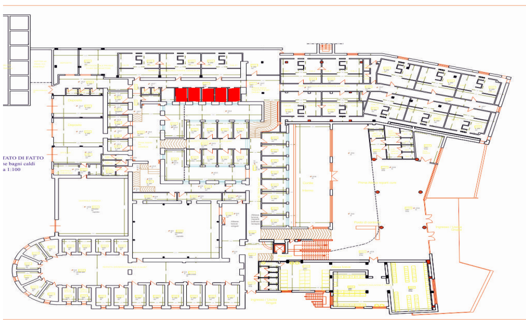

Figure 4: Massage department

There are distinct warm bath and hydromassage departments for private customers and groups. For groups there are "yellow cabins" and "green cabins" consisting of 16 warm bath/hydromassage cabins. For private customers there is an area called "horseshoe", consisting of 9 warm bath/hydromassage cabins and 7 cabins that are warm bath only. A hydromassage cabin is also suitable for a warm bath, but a warm bath cabin is not suitable for hydromassage. In addition, if one of the two departments, for example that of private customers, is busy and the other department has availability, then a private customer may access the groups department, and vice versa. For this department there are 3 operators.

5. DATA ANALYSIS

The delays occur at the entrance to the departments. The less time that the customer spends in the system, the greater his or her satisfaction. The time that the customer spends undergoing therapies in the different departments varies according to the period when the customer undertakes the treatment and then according to customer flow and the type of care. Customer wait times can last hours, and the most critical months in this case are those of September and October. A large collection of data can be used to build a simulation that helps predict these times. The goal of this study is to identify delays and create scenarios that will improve efficiency.

The simulation allows the observation of criticalities that can arise in the management of real flow, according to key factors such as the volume of traffic of customers, the type of customers (whether private or members of groups), and the type of care. Obviously, the simulation is the final stage of a project lasting several months.

Building the model is equivalent to following virtually the path made by customers to undertake treatments, studying the issues and highlighting all the possible alternatives. A visit at the spa was necessary to understand, in general terms, the approach of customers, and to gather all the information necessary for the work to proceed. The frequency with which customers arrive is a key variable of our study process, considering that the period of highest flow is the month of September when there are the greatest criticalities. After careful examination of the flow of customers, thanks not only to on-site visits, but also to a list of the thermal treatments compiled and archived each month, detailing entrances in each department, it emerged that the traffic peaks of customers during the months when they undertake treatments (i.e. from May to November) occur in September. September is a critical period, in which there can be waiting times of one or two hours for delivery of the service. We focus on the most critical month, and in particular the simulation was made for the day of September 24, the day of highest average customer flow during September. From a company archive database, we have been able to extract the time at which each person entered the various departments of the health spa, because as each customer enters a thermal department, he/she must pass a magnetic card and the entrance time is registered in the database. In this way it was possible to introduce a schedule of arrivals into the simulation model, using the quarter hour as the base unit of time.

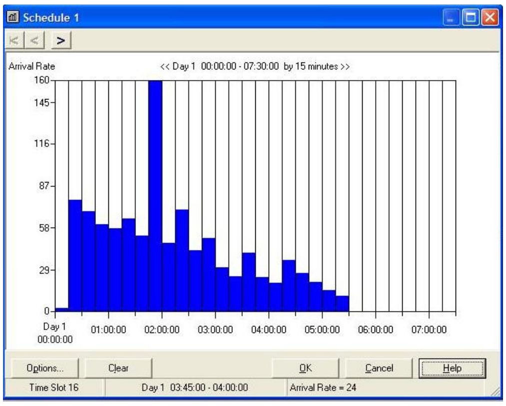

Figure 6: Arrival flow schedule

The goal of on-site sampling was to obtain service times for each type of treatment, arrival times and customer wait times. This on-site sampling was carried out for two week, in a different department each day. From these data the probability distribution of service times was obtained for each type of treatment, represented by a triangular distribution.

6. LOGICAL MODEL

The simulation model is the implementation in simulation software (Arena, in this case) of a previously-developed logical model. In general for complex service systems it is necessary to adopt the perspective of a modular system, developing the model through several sub-models.

First it is necessary to distinguish two categories of customers, customers belonging to groups that have an agreement with the company, and private customers. There are in fact two different entrances for the two types of customers. At this point we must make a further distinction by type of treatment. Both private customers and groups are sent to different departments depending on the therapy that has been assigned:

- for inhalation treatments, customers go to the inhalation department, groups on the lower floor, private customers on the upper floor;
- for warm bath or hydromassage treatments, customers go to the respective department. There are in fact two departments, one for private customers and one for groups, but if one of them is busy and the other is available, a private customer can access in the group department and vice versa;
- for massage treatment, customers go to the massage department, which is identical for both types of customers.

Each department is represented as a sub-model. We will briefly consider just one of these, the massage department. For massage treatment, customers go to the massage department, which has 5 cabins with 5 operators. As discussed above, there are two different entrances to the departments, one for groups and one for private customers. If the massage cabins are all busy, the customer remains in the queue, and the queue is formed at the control station where customer must pass the magnetic card, rather than within the massage department. There are therefore two queues, one for each type of customer. The logic of the queue in this case is for the longest queue to be dealt with first, so the condition of progress in the queue for a private customer, for example, is that one of the five cabins is free and that the private queue is longer than the group queue. At this point a distinction must be made between full massage and partial massage, which differ by the duration of treatment, about twenty minutes for the full massage and ten minutes for the partial massage. These times are not standard, but variable according to the operator and patient. The process times are defined in the block allocations, estimated through on-site sampling. Once the treatment is complete, the customer leaves the department.

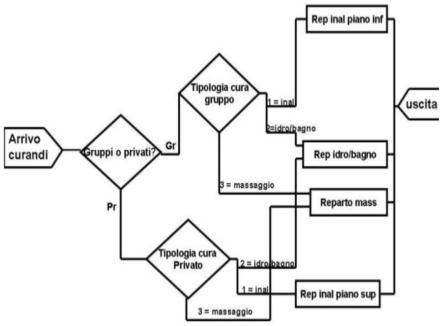

Figure 7: Overview of the logical model of the system

7. SIMULATION MODEL IN ARENA SOFTAWARE

Although the Rockwell Arena software is very powerful, it is also very easy to use. The real system is simulated through the use of blocks, linked together to form a model that reflects the real situation. For more immediate visibility, the model is developed in different modules, which represent the different steps followed by the customer in undergoing treatments.

As mentioned, the entry flow of customers is traced back to a probability distribution derived from information taken from the corporate database on the day which was chosen for simulation. The model is developed in the most general way possible, as is logical for a model that must adhere to reality.

Having extensively discussed the model from the logical point of view, we will now limit ourselves to highlighting only some salient features in order to avoid unnecessary repetition.

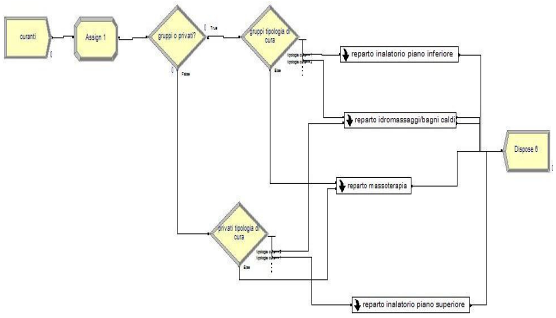

Figure 8: Graphical representation of the simulation model in Arena

Each process, aerosol, inhalation, nasal shower, hydromassage, or massage, has been represented as SEIZE - DELAY - RELEASE. In general, each process module has the ability to perform four different actions: DELAY, SEIZE, SEIZE - DELAY - RELEASE and DELAY - RELEASE. The DELAY action causes a delay of the entity without the need for a resource, the SEIZE action does not release the resource from the entity. In our case, there is an action that requires an entity to be processed and afterwards sets the entity free; SEIZE - DELAY - RELEASE is therefore the most appropriate action. By representing the processes in this way, the queues are managed autonomously by the process blocks according to a FIFO logic without having to introduce a HOLD block, a type of block specifically for the management of the queues. The service times of each process have been represented with a triangular distribution around the average. The values are those recorded during on-site sampling. In the case of the massage department and warm bath and hydromassage department, for the management of the queues we had to introduce HOLD blocks in which an expression of progress of the queue is inserted.

To define the number of runs for any simulation we need to make some statistical considerations. In fact, each simulation provides as output the average value of the variable calculated during the simulation together with the half-width of the confidence interval with which that variable was estimated. Each confidence interval is accompanied by the relative level of significance which is a measure of risk of error in estimating the value of a statistical variable. The range of values between the lower and upper limit depends on the level of probability selected and the sample size. So, although we cannot say if the real value of the variable is contained within the confidence interval estimated, the narrow confidence intervals indicate that the estimate obtained with the sample is an accurate estimate of the variable. All design decisions should be taken based on statistically significant results, which are obtained by increasing the number of runs in order to obtain an acceptable confidence interval with a significance level of 95%, provided by default by ARENA. Therefore, indicating with n_0 a minimum number of runs (at least 10) we have:

$$n = n_0 * (h_0 / h)^2$$

where h_0 is the half-width of the confidence interval corresponding to the number of replications n_0 (Half-width/Average) and h is 0.05.

We then proceed with the first simulation with ten replications. From the graph we can see that the greatest queues occur in the inhalation department.

Making the necessary calculations, we concluded that in this case we should make 260 replications for each simulation to have a significant estimate for all parameters.

Output

Output	Average	Half Width	Minimum Average	Maximum Average
C_T	13417.37	1.617,63	10637.82	17073.78
costo disservizio	8179.37	1.617,63	5399.82	11835.78
COSTO_INFR	4089.00	0,00	4089.00	4089.00
Tempo_attesa_mass_gr	0.4769	0,09	0.2783	0.6896
Tempo_attesa_mass_pr	0.5490	0,14	0.2616	0.8455
Tempo_attesa_medio_aerosol	1.0076	0,08	0.7096	1.1090
Tempo_attesa_medio_aerosolpr	0.3623	0,05	0.2111	0.4914
Tempo_attesa_medio_bagnocaldo_pr	0.01596383	0,02	0.00	0.06653352
Tempo_attesa_medio_doccianasale	0.02304063	0,00	0.01769247	0.02917486
Tempo_attesa_medio_doccianasalepr	0.02198476	0,00	0.01719767	0.02644154
Tempo_attesa_medio_idromassaggio_gruppi	0.2649	0,03	0.1775	0.3180
Tempo_attesa_medio_idromassaggiopr	0.1540	0,03	0.07444875	0.2330
Tempo_attesa_medio_inalazione	0.03902494	0,00	0.03043131	0.05007024
Tempo_attesa_medio_inalazionepr	0.05096267	0,01	0.03459430	0.07347233
tepo_attesa_medio_bagnocaldogr	0.04095608	0,04	0.00	0.1855

Figure 9: Waiting times with 10 replications

7.1. Optimization with Optquest for Arena

The ultimate goal of any mathematical model is not solely to estimate the costs, but to optimize them according to some criteria. Optimization involves defining an appropriate objective function and the related criterion; in fact we must decide if the objective is to minimize or maximize. In our model, the parameter that we want to optimize is the "Total Cost". The value of the objective function will change according to the range of inputs, or combinations of resources. Our interest is to calculate the value of the function for different inputs belonging to a defined domain, and to determine with which input values produce the minimum of the objective function. The total cost function, in this case, is a total daily cost for the simulated day, which can be seen as the sum of three contributions:

- costs of infrastructure resources (aerosol device, inhalation, hydromassage, nasal douche, warm baths, massage cabins);
- cost of human resources (number of operators for each department);
- cost of customer waiting time in different queues.

The cost of infrastructure resources is measured as the product of the unit cost of each available resource and the number of each, all multiplied by a weighting factor. The cost of human resources is measured as the sum of products of the number of operators for each department and the daily unit cost of each operator, again multiplied by a weighting factor. The third rate, that is the cost of inefficiency, has been quantified in order to give weighting to average waiting times in queues that exceed a tolerance limit set to 20 minutes. In particular, the greater the difference between average waiting time in queues, and the maximum tolerated waiting time, the higher the cost of inconvenience of the customers. The three weighting factors are different in order to produce rates with the same order of magnitude, giving slightly more weight to the inconvenience of customers waiting in queue, according to the requirements of the company.

Considering also structural constraints and the budget of costs in a year reported to the day, OptQuest for Arena provides as the best solution (after 1000 simulations, each of them with the optimal number of replications as previously calculated i.e. 260): the addition of 14 new aerosol devices for groups and 4 for private customers, 11 new inhaler devices for groups and 2 for private customers, one more nasal douche for both groups and private customers, one more massage cabin, the removal of 4 warm baths, and the addition of one operator in each department.

With this new resource configuration there is a substantial reduction in the total cost by 45% and in particular, an 11% increase of resource costs corresponds to the significant reduction of about 76% in the cost of inefficiency.

Consequently, because the cost of inefficiency is quantified so as to give weighting to average waiting time in queue for customers exceeding a tolerance limit of 20 minutes, with the new sub-optimal configuration calculated by Optquest according to the variables, constraints and objectives provided as requested by the company, we have significantly reduced the waiting times and therefore obtained a higher quality of service

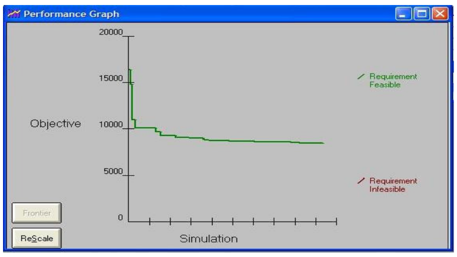

Figure 10: Performance graph

8. CONCLUSIONS

The optimal combination of resources allows minimization of the inefficiencies of service in terms of reducing queues and customer waiting times. The new configuration has lowered the cost of inefficiency in relation to the fact that the average waiting time in queues, and similarly the average number of customers in the queue are greatly reduced. Therefore we can see the importance of the simulation model, which produce considerable benefits both for the company, which can thus use the space available in a more rational way, and for customers due to the reduction of waiting times and therefore the probability that they may not be satisfied by the service. These advantages are not just in terms of space but also in terms of costs.

We believe that the model presented in this project can dynamically assist managers in their vital role, developing the optimal solution for the particular context. Although the results are satisfactory, in the future the optimization could be improved through scheduling of the available resources, so that some resources do not remain active for the entire period of the simulation, but open and close in response to queues. We can also imagine the scheduling of arrivals, creating a kind of call center system so as not to accumulate too many arrivals in one day, and consequently to reduce queues. The main advantage of using a software simulation is the ability to explore different scenarios or new methods without incurring the cost of experiments in the real system. After validating the model, changes can be made and the effects of these can be observed directly on the computer. On the other hand, the model needs to be kept up to date because the company operates in a highly dynamic context and is still evolving, so the operational scenario in a few years could be significantly different from the current one. The simulation model was developed in a very flexible way, allowing input data to be easily changed and adapted to the particular situation of reference. In this way it is possible to enter alternative configurations, change the process times in relation to the acquisition of new technologies, etc. For these reasons, the importance of organizing the system in sub-models is clear.

REFERENCES

Anagnostakis I., Dietemar B., 2001. Runway Operations Planning and Control Sequencing and Scheduling. HICSS '01 Proceedings of the 34th Annual Hawaii International Conference on System Sciences (HICSS-34)-Volume 3.

Pinedo M., 2004. Planning and Scheduling in Manufacturing and Services. Springer Series in Operations Research. Springer.

Denton B., Viapiano J., Vogl A., 2007. Optimization of surgery sequencing and scheduling decisions under uncertainty. Health care Management Science.

Ben-Daya M., Al-Fawzan, 1998. A tabu serch approach for the flow shop scheduling problem. European Journal of Operational Research 109.

Ernst A.T., Jiang H., Krishnamoorthy M., 2004. Staff scheduling and rostering: A review of applications, methods and models. European Journal of Operational Research. Volume 153, Isuue 1.

McLean A., Biles E. W., 2008. A simulation approach to the evaluation of operational costs and performance in linear shipping operations.

Proceedings of the 2008 Winter Simulation Conference.
Landeghem V.H., Bauselinck A., 2002. Reducing passenger boarding time in airplanes: A simulation based approach. European Journal of Operational Research 142.
Bruzzone A.G., Longo F., 2009. An Advanced System for Supporting the Decision Process within Large Scale Retail Stores. Simulation Transaction of the Society for Modeling & Simulation International, doi:10.1177/0037549709348801.
Bruzzone A., Signorile R. 2001. Container Terminal Planning by Using Simulation and Genetic Algorithms, Singapore Maritime & Port Journal, pp. 104-115 ISSN 0219-1555.
Bruzzone A.G., Fadda P., Fancello G., Bocca E., D'Errico G., Massei M., 2009. Virtual world and biometrics as strongholds for the development of innovative port interoperable simulators for supporting both training and R&D. Int. J. Simulation and Process Modelling, Volume 6 - Issue 1 - pp. 89 - 102 DOI 10.1504/IJSPM.2010.032661.
Revetria R., Tonelli F. Mosca R. Pozzi Cotto S. 2002. Manufacturing management Training Through Process-oriented Simulation Tool. International Journal of Simulation series ISSN 0735-9276, Volume 34 Part 1 pp. 58-64.
Revetria R., Mosca R., Giribone P., Cassettari L., Cipollina S. 2008. Theoretical Development and Applications of the MSPE Methodology in Discrete and Stochastic Simulation Models Evolving in Replicated Runs. Journal of Engineering, Computing and Architecture ISSN 1934-7197 Volume 2, Issue 1.

AUTHORS BIOGRAPHY

TERESA MURINO graduated in Mechanical Engineering, is assistant professor in the ING-IND 17, Industrial Plant System disciplinary group, in the Faculty of Engineering at University of Naples "Federico II". She teaches Manufacturing System Management, Goods and Services Production System, and Industrial Logistics at Engineering Faculty. She is also Professor at "Consorzio Nettuno". She is also peer-reviewer for Elsevier Editorials, and other journal ISI indexed.
The research activities is mainly concerned about the following topics: Simulation modelling; Maintenance strategies; Supply Chain Management models; Quick Response Manufacturing; Sustainable production processes; Location-Routing and vehicle routing Problem, Lean Service and Lean production implementation. Her email address is murino@unina.it.

ELPIDIO ROMANO graduated in Transportation Engineering, is PhD in Transportation Systems and Theory. At the moment he is a contract researcher at the Department of Materials Engineering and Operations Management of the University of Naples "Federico II" Faculty of Engineering at. He is a Tutor at "Uninettuno University". The research activities is mainly concerned about the following topics: Simulation modelling; Traffic and Transportation simulation and analysis; Maintenance strategies; Supply Chain Management models; Quick Response Manufacturing; Sustainable production processes; Location-Routing and vehicle routing Problem, Lean Service and Lean production implementation. He is author of several works presented at international conferences and issued on international journals of industrial engineering. His e-mail is elromano@unina.it.

MOSÈ GALLO is a PhD of "Production Systems and Technologies" (Department of Materials Engineering and Operations Management of the University of Naples "Federico II"). He also has a master degree in mechanical engineering from the University of Naples "Federico II". At the mo- ment he is a contract researcher at the Department of Materials Engineering and Operations Man- agement of the University of Naples "Federico II". He has several research interests and related to the design and management problems of production systems. In particular he has treated issues concerning quality, maintenance, production planning, environmental sustainability and soft com- puting techniques applied to industrial fields. He is author of several works presented at interna- tional conferences and issued on international journals of industrial engineering. His email address is mose.gallo@unina.it.